

# Robust and Efficient One-way MANOVA Tests

Stefan Van Aelst and Gert Willems

Department of Applied Mathematics and Computer Science  
Ghent University, Krijgslaan 281 S9, B-9000 Ghent, Belgium.

## Abstract

We propose robust tests as alternatives to the classical Wilks' Lambda test in one-way MANOVA. The robust tests use highly robust and efficient multi-sample multivariate S- or MM-estimators instead of the empirical covariances. The properties of several robust test statistics are compared. Under the null hypothesis, the distribution of the test statistics is proportional to a chi-square distribution. As an alternative to the asymptotic distribution, we develop a fast robust bootstrap method to estimate the distribution under the null hypothesis. We show when it is asymptotically correct to estimate the null distribution in this way and we use simulations to verify the performance of the bootstrap based tests in finite samples. We also investigate the power of the new tests, as well as their robustness against outliers. Finally, we illustrate the use of these robust test statistics on a real data example. Some additional results are provided as supplemental material.

*Keywords:* MM-estimators, Fast and Robust Bootstrap, Wilks' Lambda, Outliers.

## 1 Introduction

In one-way multivariate analysis of variance (MANOVA) the interest lies in testing the hypothesis that  $k$  groups of multivariate observations share a common mean vector. It is the direct multivariate analogue of one-way univariate ANOVA and is able to deal with possible correlations between the variables under consideration. Classical MANOVA is usually based on comparing two matrices of sums of squares and cross-products, one involving the individual group means and the other using the overall mean. The test statistic then particularly compares some scale measure between the two matrices, such as the determinant, the trace or the largest eigenvalue of the matrices (see e.g. Olson 1974).

Let  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \mathbb{R}^p$  denote the complete sample, composed of  $k$  groups  $\mathcal{X}_1, \dots, \mathcal{X}_k$  with  $k > 1$ , and let  $\Pi_j = \{i : \mathbf{x}_i \in \mathcal{X}_j\}$  denote the set of indices of the observations from  $\mathcal{X}$

belonging to  $\mathcal{X}_j$  ( $j = 1, \dots, k$ ). For each group  $j = 1, \dots, k$ , let  $\mathcal{X}_j = \{\mathbf{x}_i : i \in \Pi_j\}$  be a set of independent and identically distributed observations from the  $p$ -variate distribution  $F_j$ . The number of observations in group  $j$  is denoted by  $n_j$ , and  $F_{n_j}$  denotes the empirical distribution corresponding to  $\mathcal{X}_j$ . The total sample size is given by  $n = \sum_{j=1}^k n_j$ , with corresponding empirical distribution  $F_n = \sum_{j=1}^k \frac{n_j}{n} F_{n_j}$ . Suppose that the  $k$  distributions differ only through their location vector  $\boldsymbol{\mu}_j; j = 1, \dots, k$ . Then, the hypotheses of interest are

$$H_0 : \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_k, \quad \text{vs} \quad H_a : \boldsymbol{\mu}_l \neq \boldsymbol{\mu}_m \text{ for at least one } l \neq m.$$

The most well-known statistic for testing  $H_0$  is Wilks' Lambda which is defined as

$$\Lambda_n = \frac{|\sum_{j=1}^k \sum_{i \in \Pi_j} (\mathbf{x}_i - \bar{\mathbf{x}}_j)(\mathbf{x}_i - \bar{\mathbf{x}}_j)^t|}{|\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^t|} \quad (1)$$

where  $\bar{\mathbf{x}}_j$  denotes the sample mean of each group  $\mathcal{X}_j$  ( $j = 1, \dots, k$ ),  $\bar{\mathbf{x}}$  is the overall sample mean and  $|\mathbf{A}|$  denotes the determinant of the matrix  $\mathbf{A}$ . Wilks' Lambda is essentially a ratio of generalized variances and arises as the likelihood ratio statistic under the assumption that the group distributions  $F_j$  are multivariate normal with a common covariance matrix (see e.g. Johnson and Wichern 2002). The null hypothesis is rejected for small values of  $\Lambda_n$ . The exact null distribution is complicated, so commonly a  $\chi^2$  approximation is used.

It is widely known that statistics based on sample covariances such as  $\Lambda_n$  are very sensitive to outliers. Hence, inference based on such statistics can be severely distorted when the data is contaminated by outliers (see e.g. Todorov and Filzmoser 2010). A common approach to robustify statistical inference procedures is to replace the classical nonrobust estimates in these procedures by robust estimates (see e.g. Chervoneva and Vishnyakov 2011). The underlying idea of this plug-in principle is that the robust estimates reliably estimate the parameters of the distribution of the regular data and that these regular observations follow the classical model. Depending on the type of model deviation in the data this assumption may or may not be valid. Alternatively, a more theoretically justified robust likelihood ratio type test can be obtained by replacing the likelihood function with a robust objective function in which the effect of outliers is bounded. In both cases, the robust test statistics require robust estimators of scatter instead of sample covariance matrices. Many such robust estimators of scatter have been proposed in the literature, and some of these are nowadays available in statistical software. See e.g. Hubert et al. (2008) for a recent overview. Robust likelihood ratio test statistics that are related to our

proposals have been studied by e.g. Heritier and Ronchetti (1994), Markatou and He (1994) and Copt and Heritier (2007).

In this paper we use multivariate S-estimators (Davies 1987, Rousseeuw and Leroy 1987, Lopuhaä 1989) and the related MM-estimators (Tatsuoka and Tyler 2000). S and MM-estimators behave reasonably well in terms of both robustness and statistical efficiency and therefore are relatively popular in the robustness literature. Following He and Fung (2000), we extend the definition of multivariate S and MM-estimators to the multi-sample context, and apply them in one-way MANOVA. To find the null distribution of the corresponding likelihood ratio type test statistics, one can rely on asymptotic distributional results. However, such results may only provide a good approximation for large samples and are usually only valid under strict assumptions which are unlikely to be satisfied for data containing outliers. The bootstrap would be an obvious alternative to estimate the null distribution. However, it has two drawbacks. First, repeating many times the time-consuming algorithm to calculate robust estimates is computationally demanding. Second, the bootstrap results may not be robust because bootstrap samples potentially contain much more outliers than the original sample, which may cause the bootstrapped estimate to break down even if the original estimate was able to resist the outliers. Salibian-Barrera and Zamar (2002) introduced a fast and robust bootstrap (FRB) method for inference about the model parameters based on MM-estimates. The FRB solves both of the above problems by only calculating a fast approximation to the MM-estimates for each bootstrap sample. The FRB method has been extended further to multivariate settings based on S or MM-estimates (see e.g. Van Aelst and Willems 2005, Salibian-Barrera, Van Aelst and Willems 2006). Salibian-Barrera (2005) showed that the FRB works well in practice to estimate the p-value of robust score type test statistics but did not formally investigate consistency of the FRB in this context. However, the robust likelihood ratio type test statistics that we consider here converge to their asymptotic distribution at a higher rate than the rate of convergence of the robust parameter estimates. Therefore, the FRB approximations for the parameter estimates in each bootstrap sample are not necessarily sufficiently accurate to consistently estimate this null distribution. In Section 4 we derive a condition for the FRB that needs to be satisfied to consistently estimate the null distribution of the test statistics. We show that the condition can be satisfied for the test statistics that we consider, such that

we can avoid a time-consuming application of regular bootstrap or Monte Carlo simulation to establish their null distribution.

The rest of the paper is organized as follows. Section 2 introduces S- and MM-estimators in the multi-sample setting. In Section 3 we propose several robust likelihood ratio type test statistics based on S- or MM-estimators. In Section 4 we discuss the FRB method for obtaining the null distribution of the test statistics and in Section 5 we examine the finite-sample accuracy of the FRB by simulation. In Sections 6 and 7 we investigate the robustness and the power of our tests at finite-samples, and we compare their performance with that of the classical Wilks' Lambda test, a rank-based test and a test based on the popular Minimum Covariance Determinant (MCD) estimator. Section 8 contains a real data illustration and in Section 9 we finish with a discussion and directions for further research. Proofs of theorems are given in the supplemental material.

## 2 Multi-sample S- and MM-estimators

The estimators and test statistics that we consider can be written as functionals, which means that they depend on the random variables through their distribution function  $F$ . In the multi-group case,  $F = \sum_{j=1}^k \pi_j F_j$  where  $\pi_j$  denotes the probability mass corresponding to each group  $j$  ( $j = 1, \dots, k$ ). In the finite sample case, the statistics are functions of the empirical distribution  $F_n = \sum_{j=1}^k \frac{n_j}{n} F_{n_j}$  corresponding to the complete sample  $\mathcal{X}$ . For simplicity, we assume that the empirical ratio's  $\frac{n_j}{n}$  reliably estimate the group proportions  $\pi_j$ . If this would not be the case, then the empirical proportions can be replaced by other consistent estimates for the group proportions without further complications.

Rousseeuw and Yohai (1984) first introduced S-estimators for regression. Multivariate S-estimators for location and scatter were later investigated by Davies (1987) and studied further by Lopuhaä (1989). Consider loss functions  $\rho: [0, \infty[ \rightarrow [0, \infty[$  that satisfy the following conditions (see e.g. Salibián-Barrera *et al.* 2006):

(R1)  $\rho$  is real, symmetric, twice continuously differentiable and  $\rho(0) = 0$ .

(R2)  $\rho$  is strictly increasing on  $[0, c]$  and constant on  $[c, \infty)$  for some finite constant  $c$ .

Let  $\rho_0$  satisfy (R1) and (R2) and consider a  $p$ -dimensional random variable  $\mathbf{x}$  with distribution

$F$ . Then, the (one-sample) S-estimators of multivariate location and scatter are the solution  $(\hat{\boldsymbol{\mu}}_F^{(1)}, \hat{\boldsymbol{\Sigma}}_F^{(1)})$  that minimizes  $|\mathbf{C}|$  subject to

$$E_F \left( \rho_0 \left( [(\mathbf{x} - \mathbf{T})^t \mathbf{C}^{-1} (\mathbf{x} - \mathbf{T})]^{\frac{1}{2}} \right) \right) = b_0 \quad (2)$$

among all  $\mathbf{T} \in \mathbb{R}^p$  and  $\mathbf{C} \in \text{PDS}(p)$  where  $\text{PDS}(p)$  denotes the set of positive definite symmetric  $p \times p$  matrices. The constant  $b_0$  can be chosen to obtain consistency at an assumed underlying model distribution. In practice, one usually sets  $b_0 = E_{\Phi}[\rho_0(\|\mathbf{x}\|)]$  which ensures consistency at normal distributions. Note that the scatter estimator  $\hat{\boldsymbol{\Sigma}}_F^{(1)}$  can be decomposed into a scale part  $\hat{\sigma}_F^{(1)}$  and a shape part  $\hat{\mathbf{\Gamma}}_F^{(1)}$ , i.e.  $\hat{\boldsymbol{\Sigma}}_F^{(1)} = (\hat{\sigma}_F^{(1)})^2 \hat{\mathbf{\Gamma}}_F^{(1)}$  with  $|\hat{\mathbf{\Gamma}}_F^{(1)}| = 1$ . To keep the notation simple, in the finite sample case we use the subscript  $n$  instead of  $F_n$  in the notation of estimators and test statistics. For example, we write  $\hat{\boldsymbol{\mu}}_n^{(1)}$  instead of  $\hat{\boldsymbol{\mu}}_{F_n}^{(1)}$ .

Two-sample S-estimators were considered by He and Fung (2000). Their definition can directly be generalized to  $k$ -sample S-estimators as follows. For  $\mathbf{x} \sim F = \sum_{j=1}^k \pi_j F_j$ , the  $k$ -sample S-estimators of multivariate locations and common scatter are the solution  $(\hat{\boldsymbol{\mu}}_{1,F}^{(k)}, \dots, \hat{\boldsymbol{\mu}}_{k,F}^{(k)}, \hat{\boldsymbol{\Sigma}}_F^{(k)})$  that minimizes  $|\mathbf{C}|$  subject to

$$\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_0 \left( [(\mathbf{x} - \mathbf{T}_j)^t \mathbf{C}^{-1} (\mathbf{x} - \mathbf{T}_j)]^{\frac{1}{2}} \right) \right) = b_0 \quad (3)$$

among all  $\mathbf{T}_1, \dots, \mathbf{T}_k \in \mathbb{R}^p$  and  $\mathbf{C} \in \text{PDS}(p)$ . Similarly as above, the  $k$ -sample S-scale and S-shape estimators are denoted by  $\hat{\sigma}_F^{(k)}$  and  $\hat{\mathbf{\Gamma}}_F^{(k)}$  respectively. Note that for  $k = 1$  this definition reduces to the definition of one sample multivariate S-estimators above.

In this paper we use the class of Tukey biweight loss functions given by  $\rho(t) = \min(3t^2/c^2 - 3t^4/c^4 + t^6/c^6, 1)$  which satisfies conditions (R1) and (R2). The choice of the constant  $c$  determines the breakdown point of the S-estimators which can be up to 50%. However, tuning  $\rho_0$  involves a compromise for S-estimators, since a higher breakdown point corresponds to a lower Gaussian efficiency. Multivariate MM-estimators (Tatsuoka and Tyler 2000) avoid this trade-off by computing an efficient M-estimator, starting from the robustly estimated scale of the S-estimator.

Consider the scale  $\hat{\sigma}_F^{(k)}$  of the  $k$ -sample S-estimators in (3) and let  $\rho_1$  be a function satisfying conditions (R1) - (R2) and

$$(R3) \quad \rho_1(t) \leq \rho_0(t) \text{ for all } t \geq 0 \text{ and } \rho_1(\infty) = \rho_0(\infty).$$

Then, the  $k$ -sample MM-estimators for the multivariate locations and common shape are the solution  $(\tilde{\boldsymbol{\mu}}_{1,F}^{(k)}, \dots, \tilde{\boldsymbol{\mu}}_{k,F}^{(k)}, \tilde{\boldsymbol{\Gamma}}_F^{(k)})$  that minimizes

$$\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_1 \left( [(\mathbf{x} - \mathbf{T}_j)^t \mathbf{G}^{-1} (\mathbf{x} - \mathbf{T}_j)]^{\frac{1}{2}} / \hat{\sigma}_F^{(k)} \right) \right)$$

among all  $\mathbf{T}_1, \dots, \mathbf{T}_k \in \mathbb{R}^p$  and  $\mathbf{G} \in \text{PDS}(p)$  with  $|\mathbf{G}|=1$ . The MM-estimator for the common scatter matrix is then defined as  $\hat{\boldsymbol{\Sigma}}_F^{(k)} = (\hat{\sigma}_F^{(k)})^2 \tilde{\boldsymbol{\Gamma}}_F^{(k)}$ . Note that for the case  $k = 1$  we obtain the (one-sample) multivariate MM-estimators as defined by Tatsuoka and Tyler (2000). If condition (R3) is satisfied, then the function  $\rho_1$  can be tuned to achieve any given efficiency for  $\tilde{\boldsymbol{\mu}}_{j,F}^{(k)}$  or  $\tilde{\boldsymbol{\Gamma}}_F^{(k)}$  without affecting the breakdown point of the estimates, which is the same as that of the initial S-estimator and thus determined only by  $\rho_0$  (see e.g. Salibián-Barrera *et al.* 2006, Theorem 1). In practice, one usually fixes the constant  $c_0$  in the (biweight)  $\rho_0$ -function to obtain the maximal breakdown point, while the constant  $c_1 (> c_0)$  in  $\rho_1$  is tuned to additionally achieve a high Gaussian efficiency, e.g. 95%. Note that there is some loss of robustness in the sense that MM-estimators have a higher maximum bias than S-estimators.

Both S and MM-estimators can be written as the solution of a set of fixed-point equations which is a necessary condition for the application of the FRB. For example, the  $k$ -sample S-estimators in (3) satisfy the following system of  $k + 1$  fixed-point equations:

$$\hat{\boldsymbol{\mu}}_{j,F}^{(k)} = \left( E_{F_j} \left( \frac{\rho'_0(d_j)}{d_j} \right) \right)^{-1} \left( E_{F_j} \frac{\rho'_0(d_j)}{d_j} \mathbf{x} \right); \quad j = 1, \dots, k \quad (4)$$

$$\hat{\boldsymbol{\Sigma}}_F^{(k)} = \frac{1}{b_0} \sum_{j=1}^k \pi_j \left( E_{F_j} \left( p \frac{\rho'_0(d_j)}{d_j} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)}) (\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)})^t \right) + \left( E_{F_j} W(d_j) \right) \hat{\boldsymbol{\Sigma}}_F^{(k)} \right) \quad (5)$$

where  $d_j = d(\mathbf{x}, \hat{\boldsymbol{\mu}}_{j,F}^{(k)}, \hat{\boldsymbol{\Sigma}}_F^{(k)}) = [(\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)})^t (\hat{\boldsymbol{\Sigma}}_F^{(k)})^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)})]^{1/2}$  and  $W(t) = \rho_0(t) - \rho'_0(t)t$ .

### 3 Robust test statistics based on S- or MM-estimates

Let us first focus on test statistics based on S-estimators. An obvious, intuitively appealing way to robustify (1) is based on the plug-in principle. That is, the generalized variances in (1) are replaced by robust estimates of multivariate scale, assuming that the classical model holds for the regular data. This yields the test statistic

$$S\Lambda_F^1 := \left( \frac{|\hat{\boldsymbol{\Sigma}}_F^{(k)}|}{|\hat{\boldsymbol{\Sigma}}_F^{(1)}|} \right)^{1/2p} = \frac{\hat{\sigma}_F^{(k)}}{\hat{\sigma}_F^{(1)}}, \quad (6)$$

where the one-sample estimates are computed on the entire sample  $\mathcal{X}$ .

Alternatively, we directly compare the magnitude of the within-group residuals  $\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)}$  (for  $\mathbf{x} \sim F_j$ ) to the overall residuals  $\mathbf{x} - \hat{\boldsymbol{\mu}}_F^{(1)}$ . A direct comparison between the two types of residuals is easiest when using a common scale. In fact, we propose the following two test statistics:

$$S\Lambda_F^{2a} := \frac{\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_0 \left( [(\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)})^t (\hat{\boldsymbol{\Gamma}}_F^{(k)})^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)})]^{\frac{1}{2}} / \hat{\sigma}_F^{(k)} \right) \right)}{E_F \left( \rho_0 \left( [(\mathbf{x} - \hat{\boldsymbol{\mu}}_F^{(1)})^t (\hat{\boldsymbol{\Gamma}}_F^{(1)})^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_F^{(1)})]^{\frac{1}{2}} / \hat{\sigma}_F^{(k)} \right) \right)} \quad (7)$$

$$S\Lambda_F^{2b} := \frac{\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_0 \left( [(\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)})^t (\hat{\boldsymbol{\Sigma}}_F^{(k)})^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_{j,F}^{(k)})]^{\frac{1}{2}} \right) \right)}{E_F \left( \rho_0 \left( [(\mathbf{x} - \hat{\boldsymbol{\mu}}_F^{(1)})^t (\hat{\boldsymbol{\Sigma}}_F^{(k)})^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}_F^{(1)})]^{\frac{1}{2}} \right) \right)} \quad (8)$$

Clearly, the bounded function  $\rho_0$  guarantees that outlying observations only have a limited effect on these test statistics. The test statistic  $S\Lambda_F^{2a}$  in (7) can be justified more formally as follows. Given the  $k$ -sample S-scale  $\hat{\sigma}_F^{(k)}$ , it can be shown that

$$\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_0 \left( [(\mathbf{x} - \mathbf{T}_j)^t \mathbf{G}^{-1} (\mathbf{x} - \mathbf{T}_j)]^{\frac{1}{2}} / \hat{\sigma}_F^{(k)} \right) \right) \quad (9)$$

is minimized by  $(\mathbf{T}_1, \dots, \mathbf{T}_k, \mathbf{G}) = (\hat{\boldsymbol{\mu}}_{1,F}^{(k)}, \dots, \hat{\boldsymbol{\mu}}_{k,F}^{(k)}, \hat{\boldsymbol{\Gamma}}_F^{(k)})$ , i.e. the  $k$ -sample S-estimates. Thus, we have that  $0 < S\Lambda_F^{2a} \leq 1$  and smaller values of the statistic provide more evidence against  $H_0$ . For the test statistic  $S\Lambda_F^{2b}$  in (8) we similarly have that

$$\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_0 \left( [(\mathbf{x} - \mathbf{T}_j)^t (\hat{\boldsymbol{\Sigma}}_F^{(k)})^{-1} (\mathbf{x} - \mathbf{T}_j)]^{\frac{1}{2}} \right) \right) \quad (10)$$

is minimized by  $(\mathbf{T}_1, \dots, \mathbf{T}_k) = (\hat{\boldsymbol{\mu}}_{1,F}^{(k)}, \dots, \hat{\boldsymbol{\mu}}_{k,F}^{(k)})$ , and thus also  $0 < S\Lambda_F^{2b} \leq 1$ .

Let us now consider test statistics based on MM-estimators instead of S-estimators. Because the scale statistic used in the MM-estimator is the initial S-scale, the test statistic  $S\Lambda_F^1$  remains the same. However, it does make sense to consider the MM-analogues of statistics (7) and (8), which yields:

$$MM\Lambda_F^a := \frac{\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_1 \left( [(\mathbf{x} - \tilde{\boldsymbol{\mu}}_{j,F}^{(k)})^t (\tilde{\boldsymbol{\Gamma}}_F^{(k)})^{-1} (\mathbf{x} - \tilde{\boldsymbol{\mu}}_{j,F}^{(k)})]^{\frac{1}{2}} / \hat{\sigma}_F^{(k)} \right) \right)}{E_F \left( \rho_1 \left( [(\mathbf{x} - \tilde{\boldsymbol{\mu}}_F^{(1)})^t (\tilde{\boldsymbol{\Gamma}}_F^{(1)})^{-1} (\mathbf{x} - \tilde{\boldsymbol{\mu}}_F^{(1)})]^{\frac{1}{2}} / \hat{\sigma}_F^{(k)} \right) \right)} \quad (11)$$

$$MM\Lambda_F^b := \frac{\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_1 \left( [(\mathbf{x} - \tilde{\boldsymbol{\mu}}_{j,F}^{(k)})^t (\tilde{\boldsymbol{\Sigma}}_F^{(k)})^{-1} (\mathbf{x} - \tilde{\boldsymbol{\mu}}_{j,F}^{(k)})]^{\frac{1}{2}} \right) \right)}{E_F \left( \rho_1 \left( [(\mathbf{x} - \tilde{\boldsymbol{\mu}}_F^{(1)})^t (\tilde{\boldsymbol{\Sigma}}_F^{(k)})^{-1} (\mathbf{x} - \tilde{\boldsymbol{\mu}}_F^{(1)})]^{\frac{1}{2}} \right) \right)} \quad (12)$$

It is clear that by definition of  $k$ -sample MM-estimates we have that  $0 < MM\Lambda_F^a \leq 1$  and  $0 < MM\Lambda_F^b \leq 1$ . Obviously, small values of these test statistics can again be considered evidence against  $H_0$ .

## Asymptotic distribution

We now consider the case where the groups follow a unimodal elliptical distribution with possibly different centers, denoted by  $F_j = F_{\mu_j, \Sigma}$ . We can then take  $b_0 = E_{F_{0,I}}[\rho_0(\|\mathbf{x}\|)]$  in equations (2)-(3) where  $E_{F_{0,I}}$  denotes expectation with respect to the spherical distribution  $F_{0,I}$ . This ensures that the S and MM-functionals are Fisher-consistent at  $F = \sum_{j=1}^k \pi_j F_{\mu_j, \Sigma}$  as is shown by Proposition 1 in the supplemental material. Moreover, similarly to Lopuhaä (1989), we consider loss functions  $\rho$  with derivative  $\psi = \rho'$  that satisfy the following conditions.

(R4) The functions  $\psi'(t)$  and  $\psi(t)/t$  are bounded and almost everywhere continuous. Moreover,

$$E_{F_{0,I}}\left[\frac{\psi(\|\mathbf{x}\|)}{\|\mathbf{x}\|}\right] \neq 0, E_{F_{0,I}}[\psi'(\|\mathbf{x}\|)] > 0, \text{ and } E_{F_{0,I}}[\psi'(\|\mathbf{x}\|)\|\mathbf{x}\|^2 + (p+1)\psi(\|\mathbf{x}\|)\|\mathbf{x}\|] > 0.$$

The following result presents the asymptotic distribution of the robust test statistics under the null hypothesis for this setting.

**Theorem 1** *Consider  $F = \sum_{j=1}^k \pi_j F_j$  with  $F_1 = \dots = F_k$  belonging to the family of unimodal elliptical distributions, then the robust test statistics based on S or MM-estimators with loss functions that satisfy conditions (R1)-(R4) have the following limiting distributions:*

$$\begin{aligned} n(1 - S\Lambda_n^1) &\xrightarrow{\mathcal{D}} \frac{E_{F_{0,I}}[\psi_0^2(\|\mathbf{x}\|)]}{2p\gamma_0\beta_0} \chi_{p(k-1)}^2 \\ n(1 - S\Lambda_n^{2\cdot}) &\xrightarrow{\mathcal{D}} \frac{E_{F_{0,I}}[\psi_0^2(\|\mathbf{x}\|)]}{2pb_0\beta_0} \chi_{p(k-1)}^2 \\ n(1 - MM\Lambda_n) &\xrightarrow{\mathcal{D}} \frac{E_{F_{0,I}}[\psi_1^2(\|\mathbf{x}\|)]}{2pb_1\beta_1} \chi_{p(k-1)}^2 \end{aligned}$$

where  $\xrightarrow{\mathcal{D}}$  denotes convergence in distribution. For  $l = 0, 1$ , we write  $\psi_l = \rho'_l$  as before and

$$\begin{aligned} \gamma_0 &= E_{F_{0,I}}[\psi_0(\|\mathbf{x}\|)\|\mathbf{x}\|], \\ \beta_l &= E_{F_{0,I}}\left[\left(1 - \frac{1}{p}\right)\frac{\psi_l(\|\mathbf{x}\|)}{\|\mathbf{x}\|} + \frac{1}{p}\psi'_l(\|\mathbf{x}\|)\right], \\ b_l &= E_{F_{0,I}}[\rho_l(\|\mathbf{x}\|)]. \end{aligned}$$

Note that the asymptotic distributions of  $S\Lambda_n^{2a}$  and  $S\Lambda_n^{2b}$  coincide, as do those of  $MM\Lambda_n^a$  and  $MM\Lambda_n^b$ , which explains the notations  $S\Lambda_n^{2\cdot}$  and  $MM\Lambda_n$  in Theorem 1.



By a similar but more tedious derivation as for Theorem 1, it can be shown that under a sequence of contiguous alternatives (converging to  $H_0$  at rate  $1/\sqrt{n}$ ), the distribution of the three test statistics based on S-estimators is proportional to a noncentral  $\chi^2_{p(k-1)}$  with the same noncentrality parameter, which corresponds with the general results in Heritier and Ronchetti (1994). Hence, asymptotically these test statistics are essentially equivalent under both the null and the alternative distribution. The same holds for the two MM-estimators, but their noncentrality parameter is larger than for the test statistics based on S-estimators which shows that the increased efficiency of the MM-estimators results in a higher asymptotic power of the corresponding tests, as expected. However, this increase in power should not result in a large loss of robustness. We investigate this further in the next sections.

## Influence functions

We now study the robustness of the test statistics through their influence function, which quantifies the effect of a single observation on each statistic (Hampel et al. 1986). The influence function of a statistical functional  $\mathbf{T}$  at the distribution  $F$  is defined as

$$IF(\mathbf{y}, \mathbf{T}, F) = \lim_{\epsilon \rightarrow 0} \frac{\mathbf{T}(F_{\epsilon, \mathbf{y}}) - \mathbf{T}(F)}{\epsilon} = \frac{\partial}{\partial \epsilon} \mathbf{T}(F_{\epsilon, \mathbf{y}})|_{\epsilon=0}$$

where  $F_{\epsilon, \mathbf{y}} = (1 - \epsilon)F + \epsilon\Delta_{\mathbf{y}}$  and  $\Delta_{\mathbf{y}}$  denotes the point mass at  $\mathbf{y}$ . In our multigroup setting we consider partial influence functions (Pires and Branco 2002) assuming that the first group is contaminated. Hence, we take  $F_{\epsilon, \mathbf{y}} = \pi_1 F_{1, \epsilon, \mathbf{y}} + \sum_{j=2}^k \pi_j F_j$  with  $F_{1, \epsilon, \mathbf{y}} = (1 - \epsilon)F_1 + \epsilon\Delta_{\mathbf{y}}$ . We derive the influence function under the null hypothesis and in particular we assume that  $F_1 = \dots = F_k$ . Moreover,  $F_1$  is assumed to be a unimodal elliptical distribution  $F_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ . Due to affine invariance of the functionals it then suffices to derive the influence function at  $F_{\mathbf{0}, \mathbf{I}}$ .

The influence function has been extended to tests in Rousseeuw and Ronchetti (1981) followed by several applications to robust testing in linear models. See Chapters 3 and 7 of Hampel et al. (1986) for an overview. It turns out that the influence functions of our test statistics are exactly zero, as can be expected for asymptotically quadratic tests (see e.g. Heritier and Ronchetti 1994, Cantoni and Ronchetti 2001, Ronchetti and Trojani 2001, Wang and Qu 2007). Hence, the effect of a small amount of contamination on the test statistics is quantified by their second-order influence functions (see also Croux et al. 2008), defined as

$$IF2(\mathbf{y}, \mathbf{T}, F) = \frac{\partial^2}{\partial \epsilon^2} \mathbf{T}(F_{\epsilon, \mathbf{y}})|_{\epsilon=0}.$$

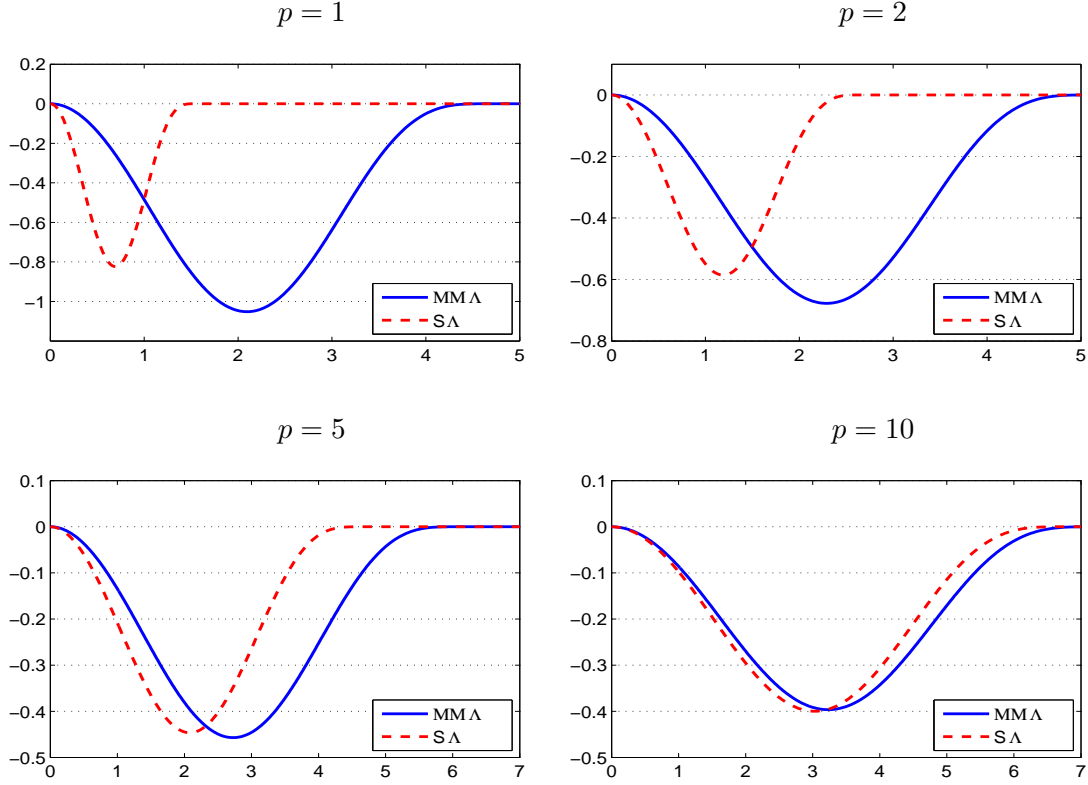


Figure 1: Standardized second-order partial influence functions of the robust test statistics at the normal model.

**Theorem 2** Consider  $F = \sum_{j=1}^k \pi_j F_j$  with  $F_1 = \dots = F_k$  having a unimodal spherical distribution  $F_{\mathbf{0}, \mathbf{I}}$ . Take  $b_0 = E_{F_{\mathbf{0}, \mathbf{I}}}[\rho_0(\|\mathbf{x}\|)]$  to ensure Fisher consistency of the  $S$  and  $MM$  functionals. Then, the second-order partial influence functions of the test statistics based on  $S$  or  $MM$ -estimators with loss functions that satisfy conditions (R1)-(R4), for contamination in the first group, are given by

$$\begin{aligned} IF2(\mathbf{y}, S\Lambda^1, F) &= \frac{-\pi_1(1-\pi_1)}{\gamma_0 \beta_0} \psi_0^2(\|\mathbf{y}\|) \\ IF2(\mathbf{y}, S\Lambda^2, F) &= \frac{-\pi_1(1-\pi_1)}{b_0 \beta_0} \psi_0^2(\|\mathbf{y}\|) \\ IF2(\mathbf{y}, MM\Lambda, F) &= \frac{-\pi_1(1-\pi_1)}{b_1 \beta_1} \psi_1^2(\|\mathbf{y}\|). \end{aligned}$$

The scales of the test statistics are essentially arbitrary. Hence, to allow comparison of the influence functions, they should first be properly standardized. Let  $R\Lambda$  denote any of these test

statistics. Then, the corresponding self-standardized influence function  $SIF2(\mathbf{y}, R\Lambda, F)$  can be obtained by dividing its second-order influence function by the asymptotic variance of the statistic  $R\Lambda$ , which can easily be obtained from Theorem 1. For S-estimates and 95% efficient MM-estimates, Figure 1 shows the self-standardized second-order influence functions of the corresponding test statistics at the normal model, for dimension  $p = 1, 2, 5$  and 10. Note that the self-standardized influence functions of all three S-based statistics coincide. Further, when the dimension increases, the relative Gaussian efficiency of the S-estimates increases and hence the difference between test statistics based on S-estimates and 95% efficient MM-estimates diminishes (up to a point when the M-step is no longer required to obtain 95% efficiency; this is true roughly for  $p \geq 15$ ).

The influence functions reflect the *redescending* behavior of the S- and MM-estimates. That is, the test statistics may be somewhat adversely affected by a small fraction of intermediate outliers, but far outliers should be harmless. These influence functions also show that the gain in efficiency of MM-estimates has some price in robustness in the sense that an outlier should be further away from the center to have no effect on the MM-based test statistics anymore.

Similarly as in e.g. Heritier and Ronchetti (1994), Cantoni and Ronchetti (2001), Ronchetti and Trojani (2001) and Wang and Qu (2007), the following expansion shows that the effect of a small amount of contamination on the asymptotic level of the test is directly controlled by the influence function of the test statistic.

**Theorem 3** Consider  $F = \sum_{j=1}^k \pi_j F_j$  with  $F_1 = \dots = F_k$  and contaminated distributions  $F_{\epsilon,n} = \pi_1 F_{1,\epsilon,n} + \sum_{j=2}^k \pi_j F_j$  with the following  $\epsilon$  contamination neighborhoods for the first group:

$$F_{1,\epsilon,n} = \left(1 - \frac{\epsilon}{\sqrt{n}}\right) F_1 + \frac{\epsilon}{\sqrt{n}} G,$$

where  $G$  is an arbitrary distribution. Let  $\Lambda$  be a statistical functional and denote  $\xi_2(\mathbf{x}, \mathbf{y}) = \frac{\partial}{\partial \epsilon_1 \partial \epsilon_2} \Lambda(\pi_1 F_{1,\epsilon_1,\mathbf{x},\epsilon_2,\mathbf{y}} + \sum_{j=2}^k \pi_j F_j)$  with  $F_{1,\epsilon_1,\mathbf{x},\epsilon_2,\mathbf{y}} = (1 - \epsilon_1 - \epsilon_2)F_1 + \epsilon_1 \Delta_{\mathbf{x}} + \epsilon_2 \Delta_{\mathbf{y}}$ . Assume that  $\xi_2(\mathbf{x}, \mathbf{y})$  is bounded and

$$n(\Lambda(F_{\epsilon,n}) - \Lambda_n) \xrightarrow{\mathcal{D}} \chi_q^2 \quad (13)$$

uniformly over the  $\epsilon$  contamination  $F_{\epsilon,n}$ . Denote the asymptotic level of the test based on  $\Lambda_n$  by  $\alpha(K)$  when the underlying distribution is  $K$ , and denote the nominal level  $\alpha(F)$  by  $\alpha_0$ . Furthermore, denote by  $H_q(\cdot, \delta)$  the cumulative distribution function of a  $\chi_q^2(\delta)$  distribution,

and by  $\eta_{1-\alpha_0}$  the  $1 - \alpha_0$  quantile of the central  $\chi_q^2$  distribution. Then, we have

$$\lim_{n \rightarrow \infty} \alpha(F_{\epsilon,n}) = \alpha_0 + \frac{\epsilon^2}{2} \kappa \iint \xi_2(\mathbf{x}, \mathbf{y}) dG(\mathbf{x}) dG(\mathbf{y}) + o(\epsilon^2), \quad (14)$$

where  $\kappa = -(\partial/\partial\delta)H_q(\eta_{1-\alpha_0}; \delta)|_{\delta=0}$ . For the special case of point-mass contamination  $G = \Delta_{\mathbf{y}}$  this reduces to

$$\lim_{n \rightarrow \infty} \alpha(F_{\epsilon,n}) = \alpha_0 + \frac{\epsilon^2}{2} \kappa IF2(\mathbf{y}, \Lambda^1, F) + o(\epsilon^2), \quad (15)$$

Note that condition (13) is stronger than requiring the existence of the influence function of the test statistic, but is guaranteed for functionals that are Fréchet differentiable (see Heritier and Ronchetti 1994, Ronchetti and Trojani, 2001) which is fulfilled for M-estimators that satisfy conditions (A.1)-(A.9) of Heritier and Ronchetti (1994).

We can apply the previous result on our test statistics  $R\Lambda$ , if we rescale them such that they asymptotically have a chi-square distribution with  $q = p(k-1)$  degrees of freedom. This leads to the following approximation for the level of the test statistics under point-mass contamination:

$$\lim_{n \rightarrow \infty} \alpha(F_{\epsilon,n}) = \alpha_0 + \epsilon^2 p(k-1) \kappa SIF2(\mathbf{y}, R\Lambda, F) + o(\epsilon^2). \quad (16)$$

This result shows that stability of the level of the test is guaranteed if the test statistic has a bounded self-standardized influence function. A similar result can be obtained to show stability of the power if the self-standardized influence function of the test statistic is bounded (see Ronchetti and Trojani, 2001).

## 4 Fast and robust bootstrap testing

The asymptotic distributions in Theorem 1 require fairly stringent assumptions. Therefore,  $p$ -values based on the asymptotic distribution may not be very accurate except if the data are approximately elliptically symmetric and for sufficiently large sample sizes. Hence, we now investigate how the fast and robust bootstrap principle can be adapted to consistently estimate the null distribution of the robust test statistics.

Consider a sample  $\mathcal{X}$  and let  $\hat{\Theta}_n \in \mathbb{R}^m$  contain all the corresponding relevant estimates (in vectorized form). For example, in case of  $k$ -sample S-estimates we set  $\hat{\Theta}_n = ((\hat{\boldsymbol{\mu}}_{1,n}^{(k)})^t \dots (\hat{\boldsymbol{\mu}}_{k,n}^{(k)})^t \text{vec}(\hat{\boldsymbol{\Sigma}}_n^{(k)})^t)^t$ . In case of MM-estimates,  $\hat{\Theta}_n$  contains the final MM-estimates as well as the initial S-estimates,

i.e.  $\hat{\Theta}_n = ((\tilde{\mu}_{1,n}^{(k)})^t \dots (\tilde{\mu}_{k,n}^{(k)})^t \text{vec}(\tilde{\Gamma}_n^{(k)})^t (\hat{\mu}_{1,n}^{(k)})^t \dots (\hat{\mu}_{k,n}^{(k)})^t \text{vec}(\hat{\Sigma}_n^{(k)})^t)^t$  Now, suppose that  $\hat{\Theta}_n$  can be represented as a solution of a set of fixed-point equations:

$$\hat{\Theta}_n = \mathbf{g}_n(\hat{\Theta}_n), \quad (17)$$

where the function  $\mathbf{g}_n : \mathbb{R}^m \rightarrow \mathbb{R}^m$  depends on the sample. For instance, for  $k$ -sample S-estimates the corresponding equations are given by (4)-(5).

Given a bootstrap sample, randomly drawn with replacement from the original sample, the recalculated robust estimates  $\hat{\Theta}_n^*$  then solve  $\hat{\Theta}_n^* = \mathbf{g}_n^*(\hat{\Theta}_n^*)$  where the function  $\mathbf{g}_n^*$  now depends on the bootstrap sample. However, the standard bootstrap is time consuming and nonrobust as explained in the introduction. Therefore, as in Salibian-Barrera *et al.* (2006, 2008), we calculate the following linear approximation to the solution in each bootstrap sample

$$\hat{\Theta}_n^{R*} := \hat{\Theta}_n + [\mathbf{I} - \nabla \mathbf{g}_n(\hat{\Theta}_n)]^{-1}(\mathbf{g}_n^*(\hat{\Theta}_n) - \hat{\Theta}_n). \quad (18)$$

A detailed discussion on the robustness and consistency of the distribution of the FRB estimates  $\hat{\Theta}_n^{R*}$  for (one-sample) S and MM-estimates can be found in Salibian-Barrera *et al.* (2006). Note that since the FRB approximations are based on first order Taylor expansions, the FRB can not be expected to have second order accuracy properties whereas the classical bootstrap does in many cases.

Let us now consider the test statistics based on S-estimates as defined in (6)-(8) or based on MM-estimates as defined in (11)-(12). We can write each of the robust test statistics as a function of all the robust estimates involved, denoted by  $\hat{\Theta}_n$  as before. For S-estimates,  $\hat{\Theta}_n$  consists of both the one-sample and  $k$ -sample S-estimates. For MM-estimates,  $\hat{\Theta}_n$  contains the one-sample and  $k$ -sample MM-estimates as well as the corresponding initial S-estimates. Hence, for each of the robust test statistics, which we again generally denote by  $R\Lambda_n$  we have that

$$R\Lambda_n = h_n(\hat{\Theta}_n) \quad (19)$$

for some real function  $h_n$ , which depends on the data. Let  $\hat{\Theta}_n^*$  again denote bootstrapped parameter estimates and let  $\hat{\Theta}_n^{R*}$  be FRB estimates as defined in (18).

The Taylor expansion that leads to (18) implies that  $\hat{\Theta}_n^{R*} = \hat{\Theta}_n^* + O_p(n^{-1})$  (Salibian-Barrera *et al.* 2006). This approximation is sufficient to consistently estimate the sample distribution of

parameter estimates  $\hat{\Theta}_n$  that converge at order  $O_p(n^{-1/2})$ . However, the null distributions of the test statistics  $h_n(\hat{\Theta}_n)$  are of order  $O_p(n^{-1})$ . Hence, the FRB can only be used to estimate the distribution of the test statistics if the following stronger condition is satisfied:

$$h_n^*(\hat{\Theta}_n^{R*}) = h_n^*(\hat{\Theta}_n^*) + o_p(n^{-1}), \quad (20)$$

where  $h_n^*$  is simply the test statistic  $h_n$  applied on a bootstrap sample.

Consider a Taylor expansion of the smooth statistic  $h_n^*$  around  $\hat{\Theta}_n^*$ ,

$$h_n^*(\hat{\Theta}_n^{R*}) = h_n^*(\hat{\Theta}_n^*) + \nabla h_n^*(\hat{\Theta}_n^*) \left( \hat{\Theta}_n^{R*} - \hat{\Theta}_n^* \right) + R_n,$$

where  $R_n$  is the remainder term and  $\nabla h_n^*(\cdot)$  is the vector of partial derivatives. Since the second order derivatives of  $h_n^*(\cdot)$  are bounded, we easily have that  $R_n = o_p(n^{-1})$ . Because the difference  $\hat{\Theta}_n^{R*} - \hat{\Theta}_n^*$  is only of order  $O_p(n^{-1})$ , to guarantee that the second term on the right hand side is of order  $o_p(n^{-1})$  such that condition (20) is satisfied, we need that

$$\nabla h_n^*(\hat{\Theta}_n^*) = o_p(1). \quad (21)$$

Hence, the partial derivatives of  $h_n(\cdot)$  should asymptotically vanish, otherwise the FRB approximations  $\hat{\Theta}_n^{R*}$  are not sufficiently accurate to consistently estimate the null distribution of the test statistics.

While it can be verified that condition (21) indeed holds for the statistics  $S\Lambda_n^2$  and  $MM\Lambda_n$  defined above, it does not hold for  $S\Lambda_n^1$ . This consistency problem can be solved by rewriting the  $S\Lambda_F^1$  statistic as follows. In the multi-sample setting, let  $S_F(\cdot)$  denote the M-estimator of multivariate scale that is minimized by the  $k$ -sample S-estimator in (3). That is, given locations  $\mathbf{T}_1, \dots, \mathbf{T}_k$  and shape matrix  $\mathbf{G}$ , the M-estimate of scale  $S_F = S_F(\mathbf{T}_1, \dots, \mathbf{T}_k, \mathbf{G})$  satisfies

$$\sum_{j=1}^k \pi_j E_{F_j} \left( \rho_0 \left( [(\mathbf{x}_i - \mathbf{T}_j)^t \mathbf{G}^{-1} (\mathbf{x}_i - \mathbf{T}_j)]^{\frac{1}{2}} / S_F \right) \right) = b_0. \quad (22)$$

Then, for the S-estimates we clearly have that  $\hat{\sigma}_F^{(k)} = S_F(\hat{\boldsymbol{\mu}}_{1,F}^{(k)}, \dots, \hat{\boldsymbol{\mu}}_{k,F}^{(k)}, \hat{\boldsymbol{\Gamma}}_F^{(k)})$  and similarly, for the one-sample case  $\hat{\sigma}_F^{(1)} = S_F(\hat{\boldsymbol{\mu}}_F^{(1)}, \dots, \hat{\boldsymbol{\mu}}_F^{(1)}, \hat{\boldsymbol{\Gamma}}_F^{(1)})$ . Therefore, the test statistic  $S\Lambda_F^1$  defined in (6) can be rewritten as

$$S\Lambda_F^1 = \frac{S_F(\hat{\boldsymbol{\mu}}_{1,F}^{(k)}, \dots, \hat{\boldsymbol{\mu}}_{k,F}^{(k)}, \hat{\boldsymbol{\Gamma}}_F^{(k)})}{S_F(\hat{\boldsymbol{\mu}}_F^{(1)}, \dots, \hat{\boldsymbol{\mu}}_F^{(1)}, \hat{\boldsymbol{\Gamma}}_F^{(1)})}. \quad (23)$$

The difference between (6) and (23) is important for the FRB where we consider the test statistics as a function of the parameter estimates involved, as in (19). In the function  $h_n(\hat{\Theta}_n)$

corresponding to  $S\Lambda_n^1$  as defined in (6), the S-estimates of scale (or scatter) are part of the argument  $\hat{\Theta}_n$ , whereas in the function  $h_n(\hat{\Theta}_n)$  that corresponds to  $S\Lambda_n^1$  as defined in (23), the argument  $\hat{\Theta}_n$  only consists of the S-estimates of the locations and shapes, but the calculation of the corresponding M-scales is now part of the function  $h_n$  itself. It can be verified that if we use definition (23) for test statistic  $S\Lambda_n^1$ , then condition (21) is satisfied again and thus the corresponding FRB estimates consistently estimate the distribution of the test statistic. A small disadvantage of definition (23) is that once we have obtained the fast FRB approximations of the locations and shapes, we still need to compute two M-estimates of scale for each bootstrap sample. Solving (22) is done by iteration and hence the bootstrap procedure loses some effectiveness regarding computation time. However, this extra cost is very limited compared to a full classical bootstrap procedure.

When using bootstrap to estimate the null distribution of a test statistic, we need to ensure that the bootstrap samples are generated in a manner that is consistent with the null hypothesis (see e.g. Hall and Wilson 1991). That is, simply drawing samples with replacement from the original data (stratified over the  $k$  samples) is likely to yield misleading results. Therefore, we first transform the data into *null data* by translating each of the  $k$  samples so that they all have the same S- or MM-location estimate. For example, for tests based on S-estimates, we construct null data as follows

$$\mathbf{x}_i^{(0)} = \mathbf{x}_i - (\hat{\boldsymbol{\mu}}_{j,n}^{(k)} - \hat{\boldsymbol{\mu}}_n^{(1)}),$$

for  $j = 1, \dots, k$  and  $i \in \Pi_j$ . Bootstrap samples are then drawn with replacement from  $\mathbf{x}_i^{(0)}; i \in \Pi_j$ , respectively for each group  $j$ . To obtain the FRB approximations to the S-estimates in each bootstrap sample, we require in (18) the S-estimates  $\hat{\Theta}_n^{(0)}$  of the transformed original sample. Due to affine equivariance the k-sample S-estimates are immediately found as  $\hat{\Sigma}_n^{(k,0)} = \hat{\Sigma}_n^{(k)}$  and  $\hat{\boldsymbol{\mu}}_{j,n}^{(k,0)} = \hat{\boldsymbol{\mu}}_n^{(1)}; j = 1, \dots, k$ , while the definition of the one-sample S-estimates implies that  $\hat{\boldsymbol{\mu}}_n^{(1,0)} = \hat{\boldsymbol{\mu}}_n^{(1)}, \hat{\Sigma}_n^{(1,0)} = \hat{\Sigma}_n^{(k)}$ . For tests based on MM-estimates, the null data is constructed similarly, that is by translating each observation by  $(\tilde{\boldsymbol{\mu}}_{j,n}^{(k)} - \tilde{\boldsymbol{\mu}}_n^{(1)})$ . In this case, to obtain  $\hat{\Theta}_n^{(0)}$ , explicitly re-computing the one-sample S and MM-estimates is required as these do not follow from equivariance properties anymore.

From now on, we assume that bootstrap samples are drawn from the null data, not from

the original data. Then, based on  $B$  bootstrap samples, the  $p$ -value of the test is estimated as

$$\hat{p} = \frac{\#\{h_n^*(\hat{\Theta}_n^{R*})[l] < h_n(\hat{\Theta}_n)\} + 1}{B + 2} \quad (24)$$

where  $h_n^*(\hat{\Theta}_n^{R*})[l]$  is the FRB recalculation of the test statistic for the  $l$ -th bootstrap sample.

The following result formally states the consistency of the FRB method in estimating the null distribution of our test statistics when condition (21) is satisfied. For this result, both  $\rho_0$  and  $\rho_1$  need to satisfy the following regularity conditions, which are taken from Salibian-Barrera *et al.* (2006) and assure consistency of the FRB for the distribution of  $\hat{\Theta}_n$ .

(A1) The following functions are bounded and almost everywhere continuous:

$$\frac{\rho'(t)}{t}, \frac{\rho''(t)}{t^2} - \frac{\rho'(t)}{t^3}, \frac{\rho'''(t)}{t^3} - 3\frac{\rho''(t)}{t^4} + 3\frac{\rho'(t)}{t^5}, \rho''(t) \text{ and } \frac{\rho'''(t)}{t}$$

(A2)  $E_F[\frac{\rho'(d_j)}{d_j}] \neq 0$  for  $j = 0, \dots, k$  where for S-estimators  $d_0 = d(\mathbf{x}, \hat{\boldsymbol{\mu}}_F^{(1)}, \hat{\boldsymbol{\Sigma}}_F^{(1)})$  and  $d_j = d(\mathbf{x}, \hat{\boldsymbol{\mu}}_{j,F}^{(k)}, \hat{\boldsymbol{\Sigma}}_F^{(k)})$  for  $j = 1, \dots, k$ . In case of MM-estimators, the S-functionals in these distances are replaced by the respective MM-functionals. Some additional conditions are needed for the function  $\rho_1$ :

(A3) The function  $\rho_1'''(t)t$  is bounded and almost everywhere continuous.

(A4)  $E_{F_j}[\frac{\rho_1'(d_j)}{d_j}(\mathbf{x} - \tilde{\boldsymbol{\mu}}_j)(\mathbf{x} - \tilde{\boldsymbol{\mu}}_j)^t]^{-1}$  exists for  $0 = 1, \dots, k$  where the distances are based on MM-functionals. Moreover, for  $j = 0$  we have that  $F_0 = F$  and  $\boldsymbol{\mu}_0 = \tilde{\boldsymbol{\mu}}_{j,F}^{(1)}$ , while  $\boldsymbol{\mu}_j = \tilde{\boldsymbol{\mu}}_{j,F}^{(k)}$  for  $j = 1, \dots, k$ .

**Theorem 4** *Let  $\hat{\Theta}_n$  be S- or MM-estimates as above, and let the underlying distribution of the data satisfy  $H_0$ . Suppose that conditions (R1)-(R3) as well as (A1) - (A4) are satisfied and that  $\hat{\Theta}_n$  is a consistent estimator. Let  $h_n(\hat{\Theta}_n)$  denote any of the test statistics (7), (8), (11), (12) or (23). Then, under the “null data” sampling scheme, conditional on the first  $n$  observations, along almost all sample sequences, the distribution of  $n(h_n^*(\hat{\Theta}_n^{R*}) - 1)$  converges weakly to the same limiting distribution as  $n(h_n(\hat{\Theta}_n) - 1)$ .*

In the following three sections, the finite-sample properties of the robust tests based on S- and MM-estimates are examined by simulation. The S-estimates are tuned to have a 50% asymptotic breakdown point, while the MM-estimates have the same breakdown point and are additionally tuned to achieve 95% efficiency at the normal distribution. The simulations are done in Matlab, using an implementation of the fast-S algorithm (Salibian-Barrera and



Yohai 2006), adapted to compute  $k$ -sample S-estimates. The MM-estimates are computed by iteratively reweighted least squares, as usual. Matlab code to calculate the robust test statistics can be obtained at <http://users.ugent.be/~svaelst/software/RobManova.html>.

## 5 Finite-sample accuracy of the FRB based tests

In this section we investigate how closely the actual Type I error rates of our tests match their nominal values in case of finite samples. For this purpose we generated data under the null hypothesis of equal means. In particular, the  $k$  samples were all drawn from the multivariate standard normal distribution  $N(\mathbf{0}, \mathbf{I}_p)$ , so  $\boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_k = \mathbf{0}$ . For  $k = 3$  groups, we considered the combinations of dimension  $p$  and sample sizes  $n_j$  given in Table 1. More results are given in the supplemental material. For each case 1000 samples were generated and the tests used FRB with  $B = 1000$  bootstrap samples, yielding a  $p$ -value estimate for each test as defined in (24).

In our graphical representations we show the simulation results of the cases according to increasing ratio  $n/p$ . This ratio expresses the relative sample size, taking the dimension into account. The order of the cases in Table 1 corresponds to the position of each case in the plots.

p	6	2	6	6	2	6	6	2	2	2	6	2	2	2	2
$n_1$	20	10	30	50	20	50	100	20	30	30	100	50	50	100	100
$n_2$	20	10	30	50	20	50	50	20	30	30	100	50	50	50	100
$n_3$	20	10	30	20	10	50	20	20	10	30	100	20	50	20	100

Table 1: Simulation settings under the null hypothesis to investigate the accuracy of the FRB tests.

The curves in Figure 2 represent the observed type I error rate if the tests would be performed on the 5% significance level. Hence, the curves correspond to the observed percentage of samples for which  $\hat{p} < 0.05$ , with cases ranging from small to large relative sample size. The left plot considers tests based on S-estimates, while the right plot considers MM-estimates. The results for other significance levels were found to be similar.

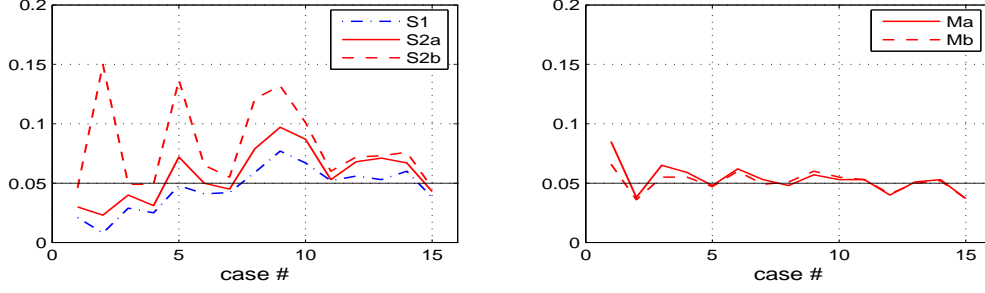


Figure 2: Observed FRB-based Type I-error rates for nominal level 0.05, various sample sizes and dimensions and for  $k = 3$  groups. The cases are ordered as indicated in Table 1.

We immediately see that the FRB based tests are more accurate for MM-estimates than for S-estimates. For the latter we find considerable deviations from the nominal 0.05 value in case of smaller samples, especially for the  $S\Lambda_n^{2b}$  test which turns out to be overly liberal. On the other hand, for MM-estimates, we see that both  $MM\Lambda_n^a$  and  $MM\Lambda_n^b$  perform nicely with Type I error rates close to 0.05. The results for  $MM\Lambda_n^b$  are slightly better than for  $MM\Lambda_n^a$ . In general, we see that the FRB based tests are more accurate for increasing relative sample size, as expected. We conclude that the FRB is a sufficiently reliable method to obtain  $p$ -values for the MM-based tests, even in small samples, while it requires larger sample sizes to obtain accurate  $p$ -values for the S-based tests.

For comparison, Figure 3 shows the type I error rates for  $k = 3$  in case the  $p$ -values are based on the asymptotic results from Theorem 1. These two plots should be compared to the plots in Figure 2. We used an estimate of the asymptotic distribution in the sense that the expected values in Theorem 1 were replaced by empirical means calculated from the sample. We see that the asymptotic approximation improves when the sample size grows, as expected, but the results are somewhat worse than for the FRB method, especially for the MM-based statistics. Similar small advantages for the FRB over the  $\chi^2$ -approximation were consistently found for other simulated underlying distributions such as long-tailed Student distributions and for outlier settings such as those considered below (results not reported here).

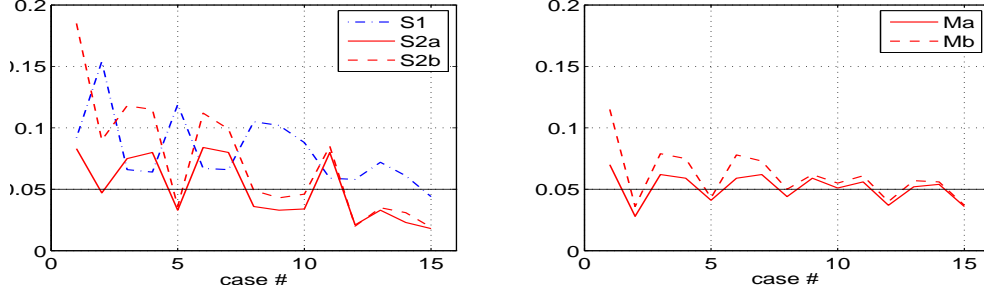


Figure 3: Observed Type I-error rates by EASV for nominal level 0.05, various sample sizes, dimensions and  $k = 3$  groups; The cases are ordered as indicated in Table 1.

## 6 Finite-sample robustness of the level

We now investigate the finite-sample robustness of the tests under the null hypothesis in comparison with the classical Wilks' Lambda test and two other tests. The first is the rank-transformed MANOVA test of Nath and Pavur (1985), which is a multivariate generalization of the better known Kruskal-Wallis test for ANOVA. It simply replaces each observation with its rank (componentwise) over all groups and then applies Wilks' Lambda. The second test (Todorov and Filzmoser, 2010) uses a test statistic based on weighted MCD estimators and obtains the null distribution by Monte Carlo simulation from a multivariate normal distribution, as implemented in the R-package `rrcov` (Todorov and Filzmoser 2009). For both the classical test and its rank-transformed version, the approximation  $-(n - 1 - (p + k)/2)\ln\Lambda_n \approx \chi_{p(k-1)}^2$  is used to obtain  $p$ -values.

The robustness is examined by generating samples as before (under the null hypothesis), but now contaminating the observations in one of the groups with probability  $\epsilon$ . In particular, groups  $\mathcal{X}_1$  to  $\mathcal{X}_{k-1}$  are generated according to  $N(\mathbf{0}, \mathbf{I}_p)$ , and group  $\mathcal{X}_k$  follows the contamination model

$$(1 - \epsilon) N(\mathbf{0}, \mathbf{I}_p) + \epsilon N(\boldsymbol{\mu}_c, \mathbf{I}_p)$$

where  $\boldsymbol{\mu}_c = d\sqrt{\chi_{p, .999}^2/p}$ , and  $d = 2, 5$  or  $10$ . We consider the setting  $k = 3$ , with sample sizes  $n_j = 20$  or  $n_j = 100$  ( $j = 1, \dots, 3$ ) and dimension  $p = 2$  or  $p = 6$ . The outlier proportion was fixed at  $\epsilon = 0.10$ . As before, 1000 samples were generated for each case.

Results given in the supplemental material show that the null distributions of the S-based and MCD-based test statistics are most resistant to the outliers in all cases. The null dis-

tributions of the MM-based test statistics are only slightly more affected by the outliers, but neither the null distributions of classical Wilks' Lambda nor its rank-transformed version are very robust.

We now investigate to what extent the Type I error rates corresponding to the estimated null distribution of the test statistics are affected by outliers. Similarly as in the previous section, Figure 4 shows observed type I error rates if the tests are performed on the 5% significance level. The rates are shown as a function of the outlier distance  $d$ , including the non-contaminated case of  $d = 0$ . The top panel of Figure 4 corresponds to the case  $n_1 = n_2 = n_3 = 20$  and  $p = 2$ , while the bottom panel shows the results for  $n_1 = n_2 = n_3 = 100$  and  $p = 6$ . The plots on the left correspond to the S-based tests, the middle plots to the MM-based tests, and the plots on the right correspond to Wilks' Lambda, its rank-transformed version and the MCD test.

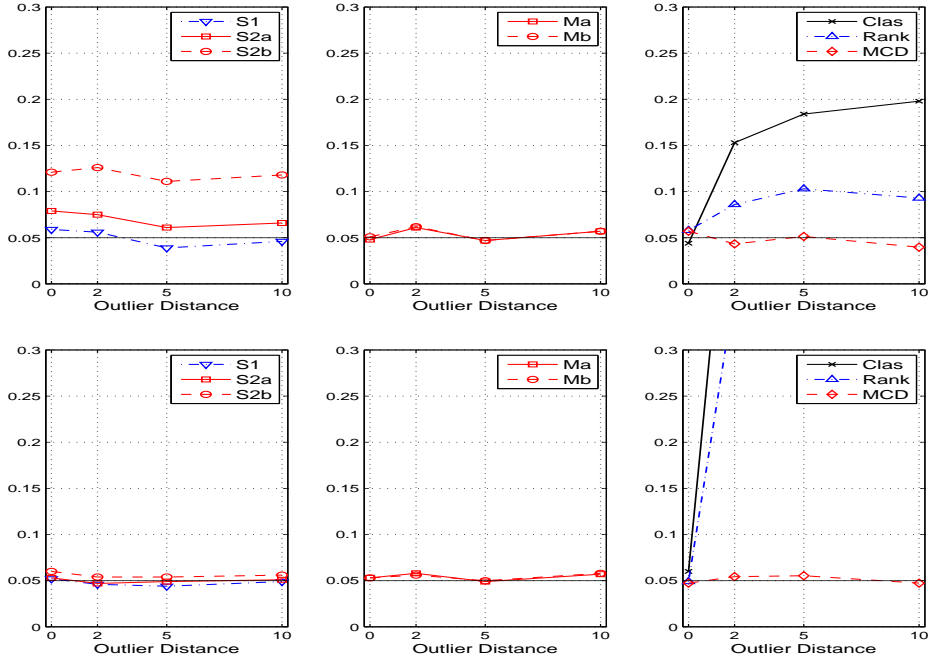


Figure 4: Observed Type I-errors for nominal level 0.05 with 10% outliers in the last group. The top panel is for the case  $n_1 = n_2 = n_3 = 20, p = 2$  while the bottom panel is for the case  $n_1 = n_2 = n_3 = 100, p = 6$ .

While the test statistics based on S-estimates have shown high resistance against the outliers, we see from Figure 4 that for  $S\Lambda_n^{2a}$  and  $S\Lambda_n^{2b}$  the  $p$ -values estimated by FRB are somewhat inaccurate for small samples. This is in accordance with the performance of FRB in

non-contaminated samples for these tests. For the MM-based tests, the Type I error rates turn out to be quite robust and close to the nominal value even though the test statistics were shown to be slightly affected. The robustness of the rank-transformed test is seen to be only slightly better than that of the classical Wilks' Lambda. Both yield very erroneous Type I error rates for the case  $n_1 = n_2 = n_3 = 100$  and  $p = 6$  (the curves fall off the plot area in the bottom right plot of Figure 4). Note that the MCD-based test performs well under contamination but recall that its null distribution is obtained by computationally demanding Monte-Carlo simulation from a multivariate normal distribution, which is the ideal case for our simulation setting. With data generated from other than a normal distribution, it can be expected that its performance is somewhat worse.

In conclusion, the tests based on MM-estimates and on the FRB method maintain their level well under contamination. In particular, they are much much more stable than classical Wilks' Lambda and the rank-transformed tests. The test statistics based on S-estimates are highly robust but the level of  $SA_n^2$  tests can be inaccurate for small samples.

## 7 Finite-sample power of the tests

Now, we compare the finite-sample power of the tests. We generate samples under the alternative hypothesis  $H_a$  as follows. Groups  $\mathcal{X}_1$  to  $\mathcal{X}_{k-1}$  are generated according to  $N(\mathbf{0}, \mathbf{I}_p)$  as before, but group  $\mathcal{X}_k$  is generated from  $N(\boldsymbol{\mu}_d, \mathbf{I}_p)$  where  $\boldsymbol{\mu}_d = (d, 0, \dots, 0)$  with  $d = 0.2, 0.5, 0.7, 1, 1.5$  or 2. We generated 1000 samples for the same dimension and sample size combinations as in the previous section.

Size-power curves given in the supplemental material show that the test statistics can differentiate well between  $H_0$  and  $H_a$ , independently of how the  $p$ -values are estimated. The classical Wilks' Lambda obviously has the highest power, but the robust MM-based statistics are almost equally powerful. They are slightly better than the rank-based statistic and also more powerful than the MCD-based and S-based test statistics.

Here, we compare the power of the tests taking into account the effect of estimating the  $p$ -values. For each of the test statistics, we estimate the corresponding  $p$ -values in the same way as in the previous section. Figure 5 shows the observed percentage of samples for which  $\hat{p} < 0.05$ , for the cases  $n_1 = n_2 = n_3 = 20$ ,  $p = 2$  and  $n_1 = n_2 = n_3 = 20$ ,  $p = 6$ . These

power curves are plotted as a function of  $d$  and include the case  $d = 0$ . The eight test statistics are distributed over the three plots as before, but the curve for the classical Wilks' Lambda is shown in each of the three plots for ease of comparison.

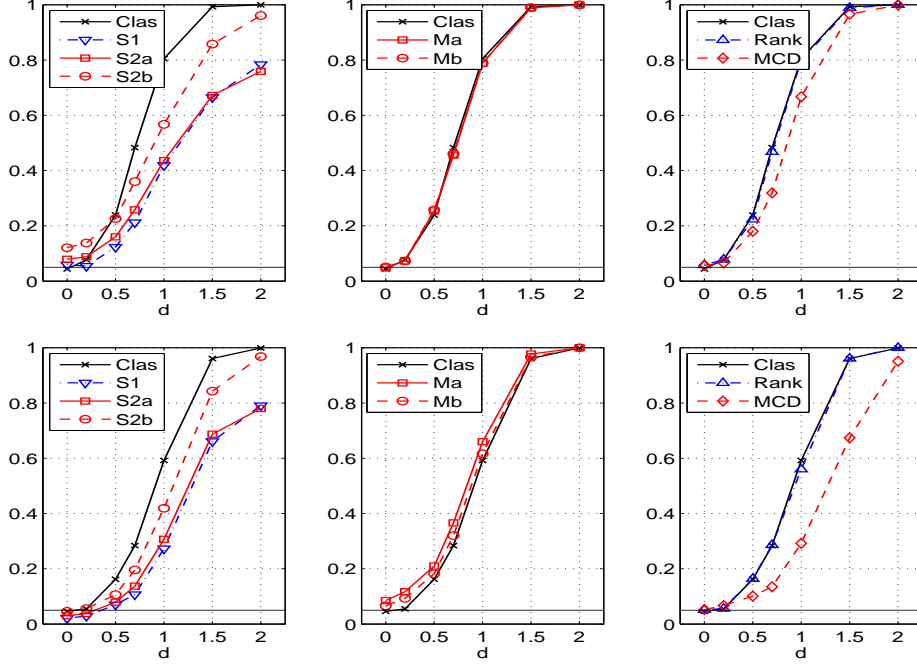


Figure 5: Power for the case  $n_1 = n_2 = n_3 = 20$ . The top panel considers the case  $p = 2$  for  $H_a : \mu_3 = (d, 0)$ . The bottom panel considers the case  $p = 6$  for  $H_a : \mu_3 = (d, 0, \dots, 0)$ .

From Figure 5 we see that both MM-based tests enjoy high power, which sometimes even exceeds that of the classical Wilks' Lambda. This is of course due to the FRB being slightly too liberal in those cases. Nevertheless we can conclude that the MM-based tests with  $p$ -values based on the FRB clearly outperform the tests based on S-estimates as well as the MCD-based test. The loss of power compared to the classical Wilks' Lambda test is very limited.

In the supplemental material, we also examine the effect of contamination on the power of the tests. From these results it can be concluded that the robust tests only show a small loss of their power with 10% of contamination in the settings of Figure 5. From the influence functions in Theorem 2 we know that the MM-based statistics may have a larger bias than the S-based robust test statistics for intermediate outliers. For larger samples, their larger bias may weaken the MM-based tests in comparison to its competitors. Simulation results with sample sizes  $n_1 = n_2 = n_3 = 100$  indeed show that in this case the S and MCD based tests enjoy better

power for small  $d$ . The S and MCD estimators have a higher variance but a lower bias than the MM-estimators for close by outliers. Especially for large  $n$  this bias effect prevails, yielding a disadvantage for the MM-based tests based on highly efficient estimators.

## 8 Example

We consider a dataset taken from the `rrcov` package (Todorov and Filzmoser 2009). The dataset contains 360 samples of different plant species collected along a 120 kilometer transect running through the city of Oslo, Norway. For each sample, the concentrations of 25 chemical elements were reported and several factors were investigated for their effect on these concentrations. Here, we consider the factor `lithology` which consists of four levels ( $k = 4$ ) and focus on three chemical elements ( $p = 3$ ): P, K and Zn. Bivariate scatter plots (not shown here) reveal that the dataset contains several outliers. Since the relative sample size (compared to the dimension) is large, we consider MM-based tests based on 90% efficient estimators to avoid a too large bias effect on the MM-based tests as discussed in the previous section. The MM-estimates indeed detect many outlying observations, as can be seen in Figure 6 which shows the weights that the observations receive in the computation of these estimates. The left panel shows the weights for the one-sample estimates, the right panel for the  $k$ -sample estimates. Note that the FRB procedure to estimate the p-value is based directly on these weights.

The classical Wilks' Lambda test yields a  $p$ -value of .704 and hence fails to reject the hypothesis that the means of the concentrations are equal among all four levels of `lithology`. On the other hand, the three tests based on S and MM-estimates with FRB based on 5000 bootstrap samples all give evidence of a significant difference in means, with  $p$ -values below .02 as can be seen from Table 2. Hence, the presence of the outliers severely affected the outcome of the classical test, leading to a wrong conclusion. On the other hand, the outliers had little effect on the robust tests which succeed in rejecting the null hypothesis.

	Clas	$SA_n^1$	$SA_n^{2a}$	$SA_n^{2b}$	$MM\Lambda_n^a$	$MM\Lambda_n^b$	MCD-test
$p$	.704	.016	.015	.014	.018	.019	.002

Table 2:  $p$ -values for the classical and robust MANOVA tests applied on the Oslo transect data to investigate the effect of `lithology` on the chemical elements P, K and Zn.

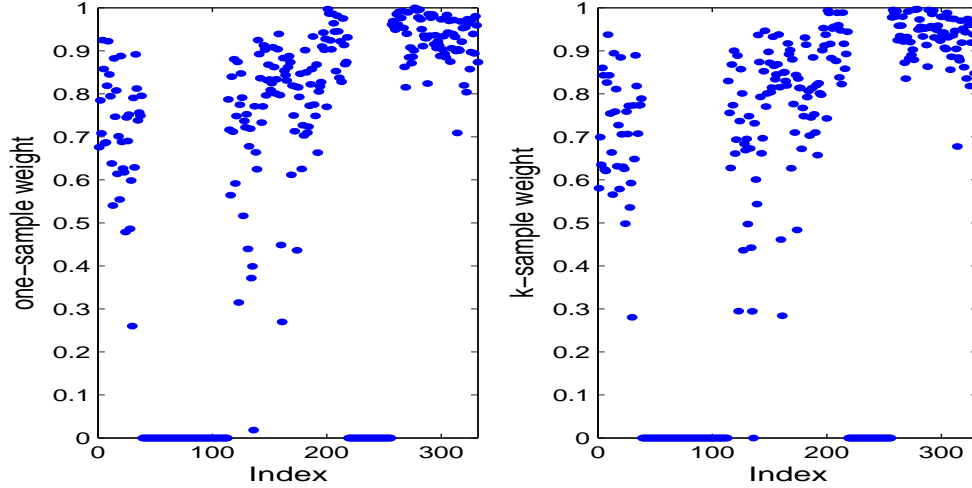


Figure 6: Oslo transect data: weights used in MM-estimator with 90% Gaussian efficiency; The top panels show the weights for the one-sample estimates and the bottom panels for the  $k$ -sample estimates.

## 9 Discussion

For small samples with relative size  $n/p$  up to 50, the MM based tests based on FRB show good performance regarding robustness of level and power. For larger relative sizes, the high efficiency of the MM-estimates, which makes them sensitive towards intermediate outliers, deteriorates the power of the tests. Therefore, we recommend to decrease the efficiency of the MM-estimates or to use the S-based test statistic  $SA_n^1$  in this case. Moreover, for relative size  $n/p$  up to 50, the FRB yields an advantage over the use of the asymptotic null distribution. For larger samples, however, this difference becomes negligible.

We have introduced the test statistics in the context of one-way MANOVA. However, the multi-group S and MM-estimates and corresponding test statistics can easily be extended to multi-way MANOVA settings. For example, for a balanced two-way MANOVA model with factors having  $k$  and  $l$  levels respectively and  $m$  observations per cell, we can consider the model

$$\mathbf{x}_{ijr} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} + \boldsymbol{\epsilon}_{ijr}; \quad i = 1, \dots, k; \quad j = 1, \dots, l; \quad r = 1, \dots, m,$$

with  $\sum_i \boldsymbol{\alpha}_i = \sum_j \boldsymbol{\beta}_j = 0$ , and  $\sum_i \boldsymbol{\gamma}_{ij} = \sum_j \boldsymbol{\gamma}_{ij} = 0$  for each  $1 \leq j \leq l$  and  $1 \leq i \leq k$  respec-



tively. The S-estimators for this model can be defined as the solution  $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\alpha}}_1, \dots, \hat{\boldsymbol{\alpha}}_k, \hat{\boldsymbol{\beta}}_1, \dots, \hat{\boldsymbol{\beta}}_l, \hat{\boldsymbol{\gamma}}_{11}, \dots, \hat{\boldsymbol{\gamma}}_{kl}, \hat{\boldsymbol{\Sigma}})$  that minimizes  $|\mathbf{C}|$  subject to

$$\sum_{i=1}^k \sum_{j=1}^l E_{F_{ij}} \left( \rho_0 \left( [(\mathbf{x} - \mathbf{m} - \mathbf{a}_i - \mathbf{b}_j - \mathbf{c}_{ij})^t \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m} - \mathbf{a}_i - \mathbf{b}_j - \mathbf{c}_{ij})]^{\frac{1}{2}} \right) \right) = b_0 \quad (25)$$

among all  $\mathbf{m}, \mathbf{a}_1, \dots, \mathbf{a}_k, \mathbf{b}_1, \dots, \mathbf{b}_l, \mathbf{c}_{11}, \dots, \mathbf{c}_{kl} \in \mathbb{R}^p$  satisfying the above model constraints and  $\mathbf{C} \in \text{PDS}(p)$ . Similarly, by setting all  $\mathbf{c}_{ij} = 0$  we can obtain the S-estimators for the two-way model without interaction effects. We can then use the test statistics (6)-(8) (with obvious modifications) to test for the existence of interaction effects. MM-estimators and MM-based tests can be constructed in a similar way. Also for MANCOVA models where extra covariates have been measured for each of the observations in the groups, multi-sample S-estimators can be defined in a manner analogous to (25).

For unbalanced multi-way MANOVA/MANCOVA problems, a multivariate regression formulation of S and MM-estimators can be used (see Van Aelst and Willems 2005). Note however, that the properties of these multivariate regression estimators have only been studied for regression models with random covariates whose distribution is non-singular. On the other hand, in unbalanced designs it often occurs that some cells contain very few or even no observations which has a large impact on the breakdown properties of the S and MM-estimators and their calculation (see e.g. Maronna and Yohai 2000, Copt and Victoria-Feser 2006). Multivariate extensions of the solutions in Maronna and Yohai (2000) that combine S-estimators and M-estimators with monotone  $\psi$  function could be useful to analyze such unbalanced MANCOVA models. This is a topic for further research.

The tests considered here focus on robustness against data contamination, but still assume that the different groups share a common covariance structure. Hence, we do not expect any improvement of our tests compared to the classical Wilks' Lambda with respect to robustness against heterogeneity of the covariance structure, but this deserves further study in the future. For Wilks' Lambda, the robustness against heterogeneity of the covariances has been extensively investigated. Rencher (1998, Section 4.2) gives an overview and concludes that only severe heterogeneity seriously affects Wilks' Lambda. An alternative test statistic, Pillai's trace statistic is even more stable in the presence of heterogeneity of the covariances. This test statistic can be written as  $\text{trace}((\boldsymbol{\Sigma}_F^{(1)} - \boldsymbol{\Sigma}_F^{(k)})/\boldsymbol{\Sigma}_F^{(1)})$ . In future research robust versions of this test statistics can be studied and compared to the robust likelihood ratio type tests introduced

here.

## 10 Supplemental Materials

**Extra details** In the supplemental file we show Fisher-consistency of  $k$ -sample S and MM-estimators and provide extra results on the finite sample performance of the tests. We also give a second real data example and proofs of the theorems.

## Acknowledgment

We would like to thank the Associate Editor and two reviewers for their helpful comments and suggestions that greatly improved the manuscript. The research of Stefan Van Aelst was supported by a grant of the Fund for Scientific Research-Flanders (FWO-Vlaanderen) and by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy).

## References

- Cantoni, E., and Ronchetti, E. (2001), “Bounded Influence for Generalized Linear Models,” *Journal of the American Statistical Association*, 96, 1022–1030.
- Chervoneva, I. and Vishnyakov, M. (2011), “Constrained S-Estimators for Linear Mixed Effects Models with Covariance Components,” *Statistics in Medicine*, 30, doi: 10.1002/sim.4169
- Copt, S. and Heritier, S. (2007), “Robust alternatives to the F-test in mixed linear models based on MM-estimates,” *Biometrics*, 63, 1045–1052.
- Copt, S. and Victoria-Feser, M.-P. (2006), “High breakdown inference in the mixed linear model,” *Journal of the American Statistical Association*, 101, 292–300.
- Croux, C., Filzmoser, P., and Joossens, K. (2008), “Classification efficiencies for robust linear discriminant analysis,” *Statistica Sinica*, 18, 581–599.
- Davies, P.L. (1987), “Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices,” *The Annals of Statistics*, 15, 1269–1292.
- Hall, P. and Wilson, S.R. (1991), “Two guidelines for bootstrap hypothesis testing,” *Biometrics*, 47, 757–762.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., and Stahel, W.A. (1986). *Robust Statistics:*

- The Approach Based on Influence Functions*, New York: Wiley.
- He, X. and Fung, W.K. (2000), “High breakdown estimation for multiple populations with applications to discriminant analysis,” *Journal of Multivariate Analysis*, 72, 151–162.
- Heritier, S. and Ronchetti, E. (1994), “Robust bounded-influence tests in general parametric models,” *Journal of the American Statistical Association*, 89, 897–904.
- Hubert, M., Rousseeuw, P.J., and Van Aelst, S. (2008), “High-breakdown robust multivariate methods,” *Statistical Science*, 23, 92–119.
- Johnson, R.A. and Wichern, D.W. (2002), *Applied multivariate statistical analysis*, Upper Saddle River: Prentice Hall.
- Lopuhaä, H. (1989), “On the relation between S-estimators and M-estimators of multivariate location and covariance,” *The Annals of Statistics*, 17, 1662–1683.
- Markatou, M. and He, X. (1994), “Bounded influence and high breakdown point testing procedures in linear models,” *Journal of the American Statistical Association*, 89, 543–549.
- Maronna, R.A. and Yohai, V.J. (2000), “Robust regression with both continuous and categorical predictors,” *Journal of Statistical Planning and Inference*, 89, 197–214.
- Nath, R. and Pavur, R. (1985), “A new statistic in the one way multivariate analysis of variance,” *Computational Statistics and Data Analysis*, 2, 297–315.
- Olson, C.L. (1974), “Comparative robustness of six tests in multivariate analysis of variance,” *Journal of the American Statistical Association*, 69, 894–908.
- Pires, A.M. and Branco, J.A. (2002), “Partial influence functions,” *Journal of Multivariate Analysis*, 83, 451–468.
- Rencher, A.C. (1998), *Multivariate Statistical Inference and Applications*, New York: Wiley-Interscience.
- Ronchetti, E., and Trojani, F. (2001), “Robust Inference with GMM Estimators,” *Journal of Econometrics*, 101, 36–79.
- Rousseeuw, P.J., and Leroy, A.M. (1987), *Robust regression and outlier detection*. New York: Wiley.
- Rousseeuw, P.J. and Ronchetti, E. (1981), “Influence Curves of General Statistics,” *Journal of Computational and Applied Mathematics*, 7, 161–166.
- Rousseeuw, P.J. and Yohai, V.J. (1984), “Robust regression by means of S-estimators,” In

- Robust and Nonlinear Time Series*. (J. Franke, W. Hardle and D. Martin, eds.). *Lecture Notes in Statistics*, 26, 256–272. Berlin: Springer-Verlag.
- Salibian-Barrera, M. (2005), “Estimating the p-values of robust tests for the linear model,” *Journal of Statistical Planning and Inference*, 128, 241–257.
- Salibian-Barrera, M., Van Aelst, S. and Willems, G. (2006), “PCA based on multivariate MM-estimators with fast and robust bootstrap,” *Journal of the American Statistical Association*, 101, 1198–1211.
- Salibian-Barrera, M., Van Aelst, S., and Willems, G. (2008), “Fast and Robust Bootstrap,” *Statistical Methods and Applications*, 17, 41–71.
- Salibian-Barrera, M. and Yohai, V. (2006), “A fast algorithm for S-regression estimates,” *Journal of Computational and Graphical Statistics*, 15, 414–427.
- Salibian-Barrera, M. and Zamar, R. H. (2002), “Bootstrapping robust estimates of regression,” *The Annals of Statistics*, 30, 556–582.
- Tatsuoka, K.S. and Tyler, D.E. (2000), “The uniqueness of S and M-functionals under non-elliptical distributions,” *The Annals of Statistics*, 28, 1219–1243.
- Todorov, V. and Filzmoser, P. (2009), “An Object-oriented framework for robust multivariate analysis,” *Journal of Statistical Software*, 32, 1–47.
- Todorov, V. and Filzmoser, P. (2010), “Robust statistic for the one-way MANOVA,” *Computational Statistics and Data Analysis*, 54, 37–48.
- Van Aelst, S. and Willems, G. (2005), “Multivariate regression S-estimators for robust estimation and inference,” *Statistica Sinica*, 15, 981–1001.
- Wang, L. and Qu, A. (2007), “Robust Tests in Regression Models With Omnibus Alternatives and Bounded Influence,” *Journal of the American Statistical Association*, 102, 347–358.